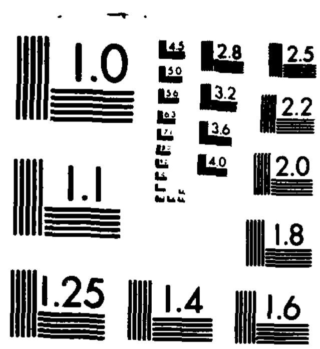
AD-A128 839 DIFFERENTIAL CHANGES WITH EXTENDED PRACTICE(U) TULANE UNIV NEW ORLEANS LA DEPT OF PSYCHOLOGY W P DUNLAP ET AL. 01 JUN 83 TR-1 N00014-80-C-0715 1/1

UNCLASSIFIED F/G 5/10 NL

END
DATE
FILMED
6-83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

AD A128839

DIFFERENTIAL CHANGES WITH EXTENDED PRACTICE

William P. Dunlap, Ina McD. Bilodeau

Tulane University

Marshall B. Jones

Pennsylvania State University College of Medicine

Final Report

June 1, 1983

This research was sponsored by:

Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research
Under Contract No. N00014-80-C-0715
Contract Authority Identification Number, NR 150-456

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report No. 1	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Differential Changes with Extended Practice	5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report 7/1/80 - 6/1/83
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): William P. Dunlap, Ina McD. Bilodeau, Marshall B. Jones	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0715
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Psychology, Tulane University, New Orleans, LA 70118	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 150-456
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs, Office of Naval Research (Code 458), Arlington, VA 22217	12. REPORT DATE: 6/1/83
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public resease; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Part I In press, Bulletin of the Psychonomic Society

Part II In press, Organizational Behavior & Human Performance

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Motor skills, video games, differential changes, extended practice, automatic processing, late appearing factors, average correlations, skills acquisition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This project was designed to answer a series of questions involving extended practice on motor skills tasks. The tasks employed were video games, and principal interest was focused on whether: (1) performance late in practice involves fewer underlying factors than early in practice; (2) video games converge or diverge differentially upon one another; and (3) factorial content of the games changes with practice. Thus far, the research has resulted in two reports. The first is a theoretical paper

(continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) addressing the issue of how best to smooth correlational data. Two methods of smoothing Pearson r's were investigated. The first was to average repeated measures in blocks then correlate block averages. In the second method all repeated measures were correlated; then the correlation coefficients were averaged in blocks. The latter approach proved much superior, resulting in greatly reduced sampling variability, and little distortion in the population correlation estimated.

The second paper examined the question of late appearing factors that should occur after a switch from controlled to automatic processing. If differential factors associated with automatic processing exist, they could only be observed after this shift occurs. Hence, any such factor would be late-appearing in the sense that it could only be identified late in practice. The second paper reports two tests of the existence of late-appearing factors. Both tests involved extended practice on five video games; the two tests were carried out in two different populations approximately one year apart. The results of the two experiments were in complete agreement. In both cases all factors, with one possible exception in the second experiment, were identified by content exclusively and not by stage of practice. The results, therefore, are negative. Other studies using other materials, other subjects, or other conditions of practice may reach different conclusions; but the studies reported in this paper offer no support for the existence of late-appearing factors.

PART 1

Average Correlations Versus Correlated Averages

Abstract

When repeated measures data are collected with small or even moderate sample sizes, correlation matrices show considerable variability. If one's primary interest is in the correlations, some means of smoothing the coefficients may be desirable. Two methods of smoothing Pearson *r*'s were investigated in the present study. The first method was to average repeated measures in blocks then correlate block averages. In the second method all repeated measures were correlated then the correlation coefficients were averaged in blocks. The latter approach proved much superior, resulting in greatly reduced sampling variability, and little distortion in the population correlation estimated.

Running Head: Average Correlations

In research involving repeated observations on the same group of subjects the correlations between the various repeated measures are often of primary experimental interest. Frequently, however, one encounters data sets where there are many repeated measures but rather few subjects. Given constraints on time and expense, the number of repeated measures and number of subjects in practical research are probably reciprocally related. In such cases the correlation matrix among repeated measures may be quite variable, frustrating attempts to make general statements about the matrix structure or attempts to decompose that structure by techniques such as factor analysis. With small to moderate sample sizes it may be of considerable benefit to smooth the correlation coefficients prior to examining the pattern of relationships. This poses the interesting question of how best to smooth these coefficients, improving their reliability while at the same time maintaining their representation of the underlying relationships.

Two possible approaches to this problem were examined by means of a computer simulation followed by some approximate mathematical predictions of the statistical properties of the two methods. The first approach examined was to average the data in blocks of measures prior to calculating the statistic of interest. This procedure has often been used in psychological research to smooth consecutive observations and we were interested in how blocking of repeated measures would affect the Pearson correlation coefficient. Blocking of data prior to computing correlations is a quite common procedure in many areas of Psychology; as examples of this pratice in the area of motor learning, see Bilodeau (1955), Fleishman (1960), or Reynolds (1952). The second, less obvious, approach examined was to find the simple correlations between all individual repeated measures, then average these correlations in blocks. A problem with the latter approach is that the sampling distribution of $\underline{r}$ is markedly skewed when the population parameter differs substantially from 0.0, thus the average $\underline{r}$ is biased toward more moderate values, under estimating the true population correlation (Kendall & Stuart, 1979). For this reason the correlations were converted via Fisher's $\underline{z}$ transformation, which almost entirely corrects skew, prior to averaging.

METHOD

The first step in investigating the properties of correlations of averages as compared to average correlations involved a simulation. The computer was programmed to generate 12 simulated "observations" having a known common population correlation for each of $\underline{N}$ subjects. Pseudo-random normal deviates (mean=0, SD=1) were generated using Box and Muller's (1958) procedure. Seeds for the normal deviates were pseudo-random rectangular fractions generated by the DEC System 20 internal function RAN (see Edgell, 1979 for statistical characteristics). First, 12 columns of independent normal data, X_{ij}, were produced; then a 13th column, W_i, was generated and was added to each of the original 12 columns to create new intercorrelated variables, Y_{ij}, by the following formula:

$$Y_{ij} = (1-\underline{r})^{1/2} X_{ij} + \underline{r}^{1/2} W_i .$$

It can be shown that the resulting columns, Y_{ij}, have population correlations equal to $\underline{r}$.

The 12 measures were then averaged in blocks of 2, 3, 4, or 6, and the correlations between block means computed. Thus, for blocks of three, the mean of Observations 1, 2, and 3 was correlated with the means of Observations 4, 5, & 6; 7, 8, & 9; and 10, 11, & 12. The program also computed the correlations between all individual observations without blocking (blocks of 1); then these correlations were averaged in blocks of 2, 3, 4, or 6. Thus, when the blocking factor was 3, the average correlation for Blocks 1 and 2 was the mean of 9 correlations, r_{14}, r_{15}, r_{16}, r_{24}, r_{25}, r_{26}, r_{34}, r_{35}, and r_{36}. Whenever correlations were averaged, they were first converted by Fisher's z transformation to avoid distortions in the means due to the skewed sampling distribution of r,

$$z = 0.5 \ln [(1+r)/(1-r)].$$

After averaging, z values were converted back to r values by the reverse transformation,

$$r = [\exp(2z)-1]/[\exp(2z)+1].$$

Five hundred data sets were generated for each combination of sample size (10, 30, or 80) and population correlation (0.2, 0.5, or 0.8). For each data set all possible interblock correlations were computed, that is, 66 correlations for "blocks" of 1 observation, 15 for blocks of 2 observations, 6 for blocks of 3 observations, 3 for blocks of 4 observations, and 1 correlation for blocks of 6 observations. The means and standard deviations (standard errors) of all correlations resulting from a given blocking, over all data sets, were then computed, again using Fisher's z. The results are shown in Table 1.

RESULTS

As can be seen in Table 1, the effect of averaging the data in blocks prior to correlating produces an increase in the block to block correlation. This increase in correlation as a function of block size is entirely consistent with the Spearman-Brown formula,

$$\underline{r}(B) = k\underline{r}/[1+(k-1)\underline{r}],$$

where $\underline{r}(B)$ is the correlation of block means, k is the number of observations per block, and $\underline{r}$ is the population correlation between observations. For example, when the correlation equals 0.2, the columns headed $\underline{r}(B)$ in Table 1 should approximately equal .20, .33, .43, .50, and .60 respectively for blocks of size 1, 2, 3, 4, and 6. As can be seen, when the sample size is 10, there is a tendency to overestimate the theoretical values, but as the sample size increases, the correlations between blocks closely approach their theoretical values. The standard errors of the $\underline{r}(B)$'s, presented in the next column of Table 1, depend only on the sample size; and since Fisher's $\underline{z}$ transformation was used, the standard errors should approximately equal $1/(\underline{N}-3)^{1/2}$. Thus, the standard errors should equal .378, .192, and .114 for samples of 10, 30, and 80, respectively. Since $\underline{r}$ is always calculated from the same number of data points, there is no gain in precision when the data are averaged in blocks prior to correlating.

Insert Table 1 about here

The next column of Table 1, headed $\bar{\underline{r}}$, shows the means of correlations averaged in blocks. As one would expect, these values simply approximate the known population correlation, with the exception that with small $\underline{N}$ there is a tendency for overestimation. This small bias is not unexpected since it is known that the mean of Fisher $\underline{z}$'s is biased in this direction (Kendall & Stuart, 1979). Thus, the averaged $\underline{r}$'s are independent of the blocking factor, and the bias toward overestimation is not great even with small sample sizes. The next column of Table 1, which shows the standard errors of the $\bar{\underline{r}}$'s is both the most interesting and the most complicated. Clearly, these standard errors decrease as a function of the blocking factor, thus there is a gain in precision by blocking. This gain in precision, however, is a function of both the blocking factor and the population correlation. When the population correlation is 0.0, the standard error should equal $1/[k(N-3)^{1/2}]$ since k^2 independent estimates of $\underline{r}$ are averaged. When the population correlation is other than 0.0, the estimates of $\underline{r}$ are themselves intercorrelated, which affects the standard errors in such a way that less precision is gained the larger the population correlation.

Steiger (1980) presents asymptotic expressions for the variance-covariance matrix between the k^2 correlations to be averaged, however, he did not directly address the issue of average correlations and their standard errors in small samples. There are k^2 variances, $2k^2(k-1)$ covariances between $\underline{r}$'s sharing one common index, and $k^2(k-1)^2$ covariances between $\underline{r}$'s with no common indices. Substituting the fact that all correlations between observations in the present problem are equal into Steiger's equations, one gets a variance of $(1-\underline{r}^2)^2$, a covariance of $\underline{r}/2(2-\underline{r}-4\underline{r}^2+3\underline{r}^3)$ for correlations sharing a common index, and a covariance of $2\underline{r}^2(1-\underline{r})^2$ for correlations with no common index. To correct for the fact that Fisher $\underline{z}$'s were used, these quantities must be divided by $(\underline{N}-3)(1-\underline{r}^2)^2$. The variance of the average $\underline{r}$, $V(\underline{r})$, may be found by pre and post multiplying the variance-covariance matrix by vectors containing $1/k^2$; thus,

$$V(\underline{r}) = \frac{(1-r^2)^2+2(k-1)r/2(2-r-4r^2+3r^3)+(k-1)^2 2r^2(1-r)^2}{k^2(N-3)(1-\underline{r}^2)^2}$$

The standard error is of course the square root of the above expression.

Using the above formula, Table 2 was constructed which gives theoretical values for the standard errors of correlations averaged in blocks under a greater variety of sample sizes and population correlations than were approximated in Table 1. Further, Table 2 presents a second column headed EN which contains the sample size required to produce the same accuracy in estimating the population correlation had sample $\underline{r}$'s not been averaged; thus Column 2 provides an estimate of the "effective sample size" or efficiency gained in averaging. These values were calculated from the following formula,

$$EN = 1/\underline{SE}^2+3,$$

which is simply the back solution from the standard error of Fisher's $\underline{z}$, where $\underline{SE}$ is the standard error and EN is the effective sample size.

Insert Table 2 about here

Comparing these theoretical values from Table 2 with the empirical values of Table 1, particularly when the sample size is smallest, one sees that the theoretical values for the standard error tend to be slightly larger, thus tend to be somewhat conservative in the case of small samples. It is clear that substantial efficiency is gained in *averaging correlations*, especially when the population correlation is small and the need for power is greatest.

DISCUSSION

First, a consideration of the limitations of the present study are in order. Our simulation covers a special case. Considerable work remains to be done by way of analyzing effective sample sizes for other specific situations, such as correlations between two sets of observations on different measures that may be differently correlated among themselves and across measures. The major point however i· ·lear. Average correlations have much smaller standard errors (muc ..rger effective N's) than correlations resulting from the commonl ·ted practice of first averaging observations.

The present paper is concerned only with the correlations of one block of observations with another block of observations, thus in calculating $\underline{r}$, k X k correlations are averaged. These derivations are not applicable to the average intercorrelations within a single block of observations, where there are k(k-1)/2 correlations to average thus different counting of the various correlation types is necessary. This latter problem is addressed by Bittner (1982). Also, the present paper does not treat the issue of a block of observations of one task correlated with a block of observations of a second task. Here, some correlations involve different observations of the same task, whereas others involve different tasks; the latter will tend to be of a lower magnitude. Preliminary investigation suggests lower efficiencies of the averaging process under these conditions (Bittner, Dunlap, & Jones, 1982).

Last it may appear that we overlooked an obvious way to improve some aspects of $\underline{r}(B)$, the coefficient obtained by correlating means of blocks of observations. The Spearman-Brown formula is easy to reverse; thus from a correlation of block means one can estimate the unblocked correlation as

$$\underline{r} = \underline{r}(B)/[k-(k-1)\underline{r}(B)].$$

Preliminary simulations were done using this back transformation on $\underline{r}(B)$, and two things were discovered. First, the mean of the back transformed $\underline{r}(B)$'s was badly biased, especially for smaller population correlations. This bias could perhaps be corrected statistically; however, ignoring the bias, the efficiencies were not nearly as great as the efficiencies of averaged correlations.

The conclusions then are rather clear. If one desires to smooth correlations, maintaining the representation of a single observation's correlation with another, while improving the coefficient's accuracy or efficiency, the average correlation is superior to the correlation of averages. It can be argued that the size of $\underline{r}(B)$, since it changes with the blocking factor, is determined by the experimenter's fancy; whereas $\underline{\bar{r}}$, since it remains almost constant across blocking, is a fairer representation of the actual experimental unit used, a single measure's correlation with another single measure. Also, $\underline{r}(B)$ fails to satisfy another primary goal of smoothing, that of improving accuracy or efficiency. One simply arrives at a bigger correlation with the same inherent instability using $\underline{r}(B)$, whereas the efficiency of $\underline{\bar{r}}$ increases dramatically, especially with lower population correlations where greatest stability is really needed. Therefore if one's interest is in the pattern of correlations across observations or a factor analysis of repeated measures data, and smoothing prior to analysis appears warranted, averaging simple correlations in blocks seems most practical.

REFERENCE NOTES

Bittner, A. C., Jr. Averaged correlations between parallel measures: Reliability estimation. Proceedings of the 8th Psychology in DoD Symposium, USAF Academy, Colorado Springs, CO., 1982.

Bittner, A. C., Jr., Dunlap, W. P., & Jones, M. B. Averaged cross-correlations with differentially stable variables: Fewer subjects required with repeated measures. Proceedings of the 26th annual meeting of the Human Factors Society, Seattle, WA, 1982.

REFERENCES

BILODEAU, E. A. Variations in knowledge of component performance and its effects on part-part and part-whole relations. *Journal of Experimental Psychology*, 1955, *50*, 215-224.

BOX, G. E. P., & MULLER, N. E. A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 1958, *29*, 610-611.

EDGELL, S. E. A statistical check of the DEC System - 10 FORTRAN pseudorandom number generator. *Behavior Research Methods & Instrumentation*, 1979, *11*, 529-530.

FLEISHMAN, E. A. Abilities at different stages of practice in rotary pursuit performance. *Journal of Experimental Psychology*, 1960, *60*, 162-171.

JONES, M. B. Differential processes in acquisition. In E. A. Bilodeau and I. McD. Bilodeau (Eds.), *Principles of skill acquisition*. New York: Academic Press, 1969.

KENDALL, M., & STUART, A. *The advanced theory of statistics*, Vol. 1, New York: Macmillan, 1979.

REYNOLDS, B. The effect of learning on the predictability of psychomotor performance. *Journal of Experimental Psychology*, 1952, *44*, 189-198.

STEIGER, J. H. Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 1980, *87*, 245-251.

Table 1. Correlations Between Block Means, r(B), and Correlations Averaged in Blocks, r, with Respective Standard Errors for Population Correlations of 0.2, 0.5, and 0.8, and Sample Sizes (N) Equal 10, 30, and 80.

Population	Blocking	N = 10				N = 30				N = 80			
Correlation	Factor	r(B)	SE	r	SE	r(B)	SE	r	SE	r(B)	SE	r	SE
	1	.209	.376	.209	.376	.201	.194	.201	.194	.199	.114	.199	.114
	2	.352	.376	.210	.221	.335	.193	.201	.115	.331	.112	.199	.066
r=0.2	3	.446	.378	.210	.173	.430	.194	.201	.090	.428	.113	.199	.052
	4	.524	.379	.211	.148	.503	.194	.201	.078	.498	.111	.199	.044
	6	.626	.378	.210	.128	.606	.188	.201	.065	.596	.114	.198	.037
	1	.515	.373	.515	.373	.508	.192	.508	.192	.500	.114	.500	.114
	2	.682	.372	.515	.259	.673	.192	.508	.135	.667	.115	.500	.080
r=0.5	3	.765	.375	.514	.224	.758	.187	.508	.116	.750	.113	.500	.069
	4	.812	.373	.514	.208	.806	.192	.508	.109	.800	.115	.500	.064
	6	.867	.374	.514	.193	.862	.185	.508	.101	.857	.114	.500	.060
	1	.821	.369	.821	.369	.808	.190	.808	.190	.803	.114	.803	.114
	2	.902	.373	.821	.286	.893	.188	.808	.148	.891	.113	.803	.090
r=0.8	3	.933	.364	.821	.261	.927	.184	.808	.136	.925	.115	.803	.083
	4	.948	.376	.820	.252	.944	.184	.808	.130	.943	.116	.803	.080
	6	.965	.362	.821	.240	.961	.183	.808	.125	.961	.112	.803	.077

Table 2. Standard Errors and Effective Sample Sizes for Correlations Averaged Over Blocks of Trials.

Sample Size	Blocking Factor	r=0.0 SE	r=0.0 EN	r=0.1 SE	r=0.1 EN	r=0.2 SE	r=0.2 EN	r=0.4 SE	r=0.4 EN	r=0.6 SE	r=0.6 EN	r=0.8 SE	r=0.8 EN	r=0.9 SE	r=0.9 EN
	1	.378	10	.378	10	.378	10	.378	10	.378	10	.378	10	.378	10
	2	.189	31	.208	26	.225	23	.255	18	.279	16	.298	14	.306	14
N=10	3	.126	66	.152	47	.176	35	.217	24	.249	19	.275	16	.286	15
	4	.094	115	.124	68	.152	46	.199	28	.235	21	.264	17	.276	16
	6	.063	255	.097	110	.129	63	.182	33	.223	23	.254	18	.268	17
	1	.243	20	.243	20	.243	20	.243	20	.243	20	.243	20	.243	20
	2	.121	71	.133	59	.144	51	.163	40	.179	34	.191	30	.196	29
N=20	3	.081	156	.097	109	.113	82	.139	55	.160	42	.176	35	.183	33
	4	.061	275	.079	161	.097	108	.128	64	.151	47	.170	38	.177	35
	6	.040	615	.062	262	.083	149	.117	76	.143	52	.163	41	.172	37
	1	.192	30	.192	30	.192	30	.192	30	.192	30	.192	30	.192	30
	2	.096	111	.106	93	.115	79	.130	62	.142	53	.152	47	.156	44
N=30	3	.064	246	.077	171	.089	128	.110	85	.127	65	.140	54	.145	50
	4	.048	435	.063	254	.077	170	.101	101	.120	73	.134	58	.141	54
	6	.032	975	.049	414	.066	235	.093	119	.113	81	.130	63	.136	57
	1	.146	50	.146	50	.146	50	.146	50	.146	50	.146	50	.146	50
	2	.073	191	.080	159	.087	136	.098	107	.107	90	.115	79	.118	75
N=50	3	.049	426	.058	295	.068	221	.084	146	.096	111	.106	92	.110	85
	4	.036	755	.048	440	.059	294	.077	173	.091	124	.102	99	.107	91
	6	.024	1695	.037	719	.050	406	.070	206	.086	139	.098	107	.103	97
	1	.102	100	.102	100	.102	100	.102	100	.102	100	.102	100	.102	100
	2	.051	391	.056	325	.060	277	.068	217	.075	182	.080	159	.082	151
N=100	3	.034	876	.041	607	.047	452	.058	298	.067	226	.074	187	.077	173
	4	.025	1555	.033	906	.041	604	.053	353	.063	253	.071	202	.074	184
	6	.017	3495	.026	1480	.035	836	.049	421	.060	283	.068	217	.072	196

PART 2

FACTORS APPEARING LATE IN PRACTICE

ABSTRACT

With extended practice on a task a shift seems to occur from controlled to automatic processing. If differential factors associated with automatic processing exist, they could only be observed after this shift occurs. Hence, any such factor would be late-appearing, in the sense that it could only be identified late in practice. The present paper reports two tests of the existence of late-appearing factors. Both tests involved extended practice on five video games; the two tests were carried out in two different populations approximately one year apart. The results of the two experiments were in complete agreement. In both cases all factors, with one possible exception in the second experiment, were identified by content exclusively and not by stage of practice. The results, therefore, are negative. Other studies using other materials, other subjects, or other conditions of practice may reach different conclusions; but the studies reported in this paper offer no support for the existence of late-appearing factors.

In recent years much attention has been focused on the growth of automatic information processing with practice (Schneider and Shiffrin, 1977; Dreyfus & Dreyfus, 1979; Jarrett, 1979). A task which requires active control early in practice may become automatic later on; the task becomes routine and the skilled performer can attend to other matters. Individual differences, however, are found in all behaviors; therefore, some people should become better automatic processors than other people. If so and if a shift to automatic processing occurs late in practice, one would expect some general abilities to emerge late in practice that could not be identified earlier on.

Factors appearing late in practice could have major practical importance. It is helpful to know how well trainees are likely to do early in training; generally speaking, however, it is more important to predict how well they will perform in operations, when training is finished. Factors associated with automatic processing might well serve the latter purpose and, if so, would contribute considerably to the predictive validity of many selection batteries.

To pursue this possibility, the formal criteria for recognizing a late-appearing factor need to be specified. Suppose that N subjects are given extended practice on k tasks and performance is scored at three points (or over three sets of consecutive trials) early, midway, and late in practice and the resulting 3k measures factor analyzed. If factor j is late-appearing, it will correlate most strongly with different tasks late in practice and less strongly with those same or other tasks earlier on. In all other cases factor j is not <u>not</u> late-appearing. If, for example, factor j correlates strongly with tasks A and B late in practice and with task C early and midway in practice, factor j is not late-appearing because it can be identified (from task C) early in practice. The only reservation is the possibility that the subjects have practiced task C or similar tasks and reached asymptotic levels before the experiment begins. In such a case all practice on task C would be late and factor $\underline{j}$ might indeed be late-appearing. If, however, the tasks under study show conventional learning curves with practice, then factor $\underline{j}$ can be late appearing only if it correlates strongly with late stages of practice exclusively. It need not correlate strongly with late practice on all tasks, but it must <u>not</u> correlate with any task midway or early in practice, at least not as strongly as it does with some tasks late in practice. The polar opposite of a late-appearing factor is a factor which correlates most strongly with one task at <u>all</u> stages of practice and less strongly with other tasks at any stage of practice. Such a factor is identified by task-content exclusively, without qualification as to stage of practice.

The hypothesis of late-appearing factors requires the same subjects to practice at least two and preferably several tasks. Only a few studies in the psychological literature meet this description (Duncanson, 1964; Gundlach, 1926; Horn, 1972; Lewis, McAllister, & Bechtoldt, 1953; Stake, 1961; Woodrow, 1946), and none was analyzed with a view to the possibility that factors might emerge with practice that could not be identified earlier on. The present study is the first to do so.

Two closely similar experiments were carried out; they will be presented separately.

EXPERIMENT 1

<u>Subjects and Procedures</u>

Eighteen Navy enlisted men between the ages of 19 and 24 and with 20/20 corrected vision served as subjects. Most of the subjects had participated in previous studies and been exposed to critical tracking, two-dimensional compensatory tracking, and dual laboratory tracking tasks. No subject, however, had been previously exposed to laboratory pursuit tracking tasks or to microcomputer-television games.[1] The subjects were fit and motivated. They received extra compensation for serving as subjects in a fully approved research and development program that meets or exceeds the criteria set down by the Navy concerning the protection of subjects in Secretary of the Navy Instruction 3900.6E (Thomas, Majeski, Ewing, and Gilbert, 1978).

The subjects practiced five tasks one session a day for 15 consecutive working days. The five tasks were all microcomputer-television games manufactured by Atari, and all were practiced in the same order by all subjects. Air Combat Maneuvering (ACM) was practiced first, followed by Breakout and Surround taken concurrently, followed by Race Car and Slalom also taken concurrently. Table 1 presents further detail on each task, including trial length, number of trials per day, and score.

Results

Means and standard deviations. Means and standard deviations for these five tasks on each day of practice have already been reported (Jones, Kennedy, & Bittner, 1981; Kennedy, Bittner, Harbeson, & Jones, 1982). All five show sharp increases early in practice followed by smaller and smaller increases with continued practice; in short, all five tasks show conventional learning curves.

Correlations. With reference to stages of practice in this experiment, "early" is defined as Days 1-5, "midway" as Days 6-10, and "late" as Days 11-15. Between any two such stages, whether of the same task or between two tasks, there are 25 correlations. These 25 correlations were averaged to obtain the figures that appear in Table 2. The value of .30, for example, which appears in the upper right-hand corner of the matrix is the average of the 25 correlations between Days 1-5 on ACM and Days 11-15 on Surround.

Averaging was carried out by first transforming each one of the 25 correlations by Fisher's _z_, averaging the z-transforms, and then transforming the average _z_ back to _r_. The standard error of this average, it should be noted, is much less than the value for a single z-transformed correlation. Simulation study of this question indicates that the effective N for the average correlation varies between 32 and 88 depending on the magnitude of the correlation (.90 to .30) as opposed to 18 regardless of magnitude (Dunlap, Jones, & Bittner, submitted).

The alternative to averaging correlations would be to average trials. The sampling error associated with a z-transformed correlation between average trials is the same as for one between single trials, hence much larger than the error associated with an average correlation.

In addition, the average correlation estimates the average between-trial correlation regardless of how these trials are grouped, whether in fives, as in the present case, threes, or what have you. The correlation between average trials, however, increases as the number of trials being averaged increases and is, therefore, partly dependent on an arbitrary consideration, namely, the way trials are grouped.

Principal factors. To obtain the principal factors for the correlations in Table 2 communalities were initially estimated as the squared multiple correlation between the variable in question and the remaining 14 variables. Factoring was continued until all remaining factors had eigenvalues of less than unity. This criterion yielded three factors. The loadings on these three factors were then used to estimate a second set of communalities and the matrix factored again (three factors). After six iterations the communalities converged. The resulting factor loadings appear in Table 3.

The three largest correlations for each factor in Table 3 are marked with an asterisk. The two largest, .88 and .83, are for stages late and midway in practice on Breakout. The next largest, .79, is for late practice on ACM. The next two, .76 and.75, are for early practice on ACM and Race Car. It does not seem, therefore, that Factor 1 can be regarded as late-appearing.

Factors 2 and 3 are not well identified at any stage of practice. To qualify as late-appearing a factor must be well identified late in practice and much less so earlier on. None of the principal factors meet this description, but few psychologists would expect them to. If late-appearing factors exist, one would expect to find them only after rotation to some sort of simple structure.

Rotated factors. Table 4 contains the correlations of each task with the varimax rotated factors. The results are unequivocal. All three factors exemplify what was described earlier as the "polar opposite" of a late appearing factor. The three largest correlations are all with the same task and all three are considerably larger than the correlations with any other tasks.

Table 5 contains the correlations of each task with the quartimax rotated factors. Again the results are unequivocal. Factor 2 is not well identified, but Factors 1 and 3 are both clearly controlled by task content.

Table 6 contains the correlations of each task with the direct oblimin, obliquely rotated factors. Again, all three factors are polar opposites of a late-appearing factor.

EXPERIMENT 2

Subjects and Procedures

The subjects were 63 male students taking introductory psychology classes at Tulane University. They were obtained in the following manner. All students in the class were asked to complete a one-page questionnaire regarding previous experience with video games. The experimenter then called for male volunteers with no more than two hours of experience with Atari home video games. (The questionnaires were checked and volunteers with more than the two-hour limit were excused from the experiment.) Volunteers were informed that the experiment would require two hours of paper-and-pencil testing (not reported in this study) and 12 hours of video testing, that they could withdraw at any time, that taking part would satisfy the instructor's requirements for research participation, and that volunteers who completed the experiment would be paid $40 each.

Video testing involved five video tasks: ACM, Breakout, Race Car, Slalom, and Antiaircraft. The last is game #1 on cassette CX-2602; each game lasts 2 min and 16 sec and is played with a joystick controller and a button for firing at targets.[2] The five games were located in as many booths in the experimental room, each one connected to a Zenith 19-in black-and-white television set. The booths were separated at the sides only by Celotex panels. The subject sat 1.5 m from the television set directly in front of a 51x62 cm table 64 cm in height. The Atari console and the subject's data sheet also rested on the table top.

Testing took place in 12 sessions, each one lasting approximately 50 min. In each session all five games were played for 9.5 min each, leaving half a minute to change seats for the next game. In all, each subject had 12x9.5 = 114 min of practice on each game. Since each game was played on a separate television set in a separate booth, a maximum of five subjects could be tested simultaneously. The order of testing for the five games was determined by the assignment of subjects to the rows of a random 5x5 latin square for each experimental session. Since each letter (representing one of the five games) appeared once in each column (representing a 9.5 min practice interval), no two subjects were assigned to play the same game at the same time. A subject was allowed no more than two sessions on a given day and was required to complete the 12 sessions within six weeks. The order of games for a given session was written across the top of the data sheet used for that session. The games were coded by the letters A through E which also appeared in large block letters above the appropriate booth. The scores each subject achieved were recorded by the subject in a vertical column labelled trials under the game's letter. At least one experimenter was present behind the subjects and verified one or more entries per player per game in a session. The subject was responsible for starting a game at the RESET signal and restarting the game after each recorded score, using the reset button at the far right of the Atari console. The game select button next to the reset button was blocked to prevent the subject from inadvertently shifting to another game. If a subject was in the middle of a game at the STOP signal, he recorded the incomplete score, then moved to the next game booth.

For the four timed games (ACM, Race Car, Slalom, and Antiaircraft) the subject's score for a session was the average of games played excluding the first and last game or part-game. For Breakout, which lasts a variable length of time, the subject's score for a session was the total number of points accumulated in the 9.5 min.

Results

Means and standard deviations. Means and standard deviations for ACM, Race Car, and Slalom, shown in Table 7, were similar to the corresponding results in Experiment 1. Since Breakout was scored by session rather than by game and a session usually included more than one game, the scores on Breakout were larger in Experiment 2 than in Experiment 1, but the learning curves were much the same. Like the other four games, Antiaircraft followed a conventional learning curve, with the mean approximately doubling from the first to the twelfth session.

Correlations. In Experiment 2 the 12 practice sessions were grouped by threes into four stages of practice. Hence, the correlations in Table 8 are averages of nine intersession correlations. Averaging was done in the same manner as in Experiment 1, that is, by averaging z-transforms rather than raw correlations.

Principal factors. Table 9 contains the principal-factor loadings for Experiment 2. The correlations were factored following the same procedures as in Experiment 1, this time resulting in four factors. Factors 2, 3, and 4 are not well identified, only two of the loadings being as high as .50. The first factor has its highest four correlations with Stages 3 and 4 on Race Car, Stage 4 on Antiaircraft, and Stage 1 on ACM. The last result excludes Factor 1 as late appearing. In addition, all loadings on Factor 1 are relatively high and distributed over a narrow range.

Rotated factors. The varimax-rotated factor loadings appear in Table 10. Three of the four factors are controlled exclusively by content, the four highest correlations all being with the same task. Factor 2 has its strongest correlation with Breakout and Race Car early in practice.

The quartimax loadings appear in Table 11. Factors 1, 3, and 4 show the now-familiar clustering of the four highest correlations with the same task. Factor 2, however, meets the requirements for a late-appearing factor. The four highest correlations are with Stages 3 and 4 exclusively and involve three tasks; furthermore, the correlations are substantial. Nevertheless, a claim that Factor 2 is late appearing seems unwarranted. Factor 2 was not late appearing in the varimax rotation, and in the quartimax rotation Stage 1 on Breakout has the fifth largest correlation with Factor 2 and a close fifth too.

The correlations in Table 12 with the obliquely rotated factors cast further doubt on Factor 2 as late appearing. Here the largest correlations are with Stages 1 and 3 on Breakout. Overall, it appears that Factor 2 correlates strongly with both Breakout and Race Car but, taking all three rotations into account, as strongly with Breakout early in practice as later on. The main difference between Factor 2 and the other three factors is that the latter are each defined by one task (ACM, Race Car, or Antiaircraft), whereas Factor 2 appears to be defined by two tasks (Breakout and Race Car).

DISCUSSION

All of the tasks studied in this report demand less attention once they stabilize than they do earlier on. Stabilization is achieved when the individual learning curves become parallel except for random perturbations and the mean learning curve is increasing slowly if at all (Jones, Kennedy, and Bittner, 1981). When this point is reached, a subject's performance is no longer changing appreciably either absolutely or relative to the performance of other individuals. In one of the first studies to use video games for psychological experimentation Rebert and Low (1978) remarked that "Excellent play also appears to involve the automatization of performance, as 300-400 hits can be achieved during simultaneous conversation." Rebert and Low were commenting on Pong, not one of our six games but one we have studied elsewhere (Note 1). Our experience is the same as Rebert and Low's, not only for Pong but also for the games studied here. All of these tasks stabilize with much less practice than the amounts given in our two studies (Note 1: Jones, Kennedy, and Bittner, 1981; Kennedy, Bittner, Harbeson, and Jones, 1982). Once they do stabilize, moreover, a reserve for carrying out other functions not only becomes available but may present problems for some uses of these tasks (Note 2). The basis of automatization, unfortunately, is not as clear as the fact itself. One cannot say, for example, that automatization depends on consistent mapping, as one can about the tasks studied by Schneider and Shiffrin (1977). The idea of consistent mapping does not apply to the tasks studied in this report or, if it does, can be brought to bear only after prolonged, detailed, and extremely difficult analysis.

The present studies are not conclusive, of course. All of the tasks studied in this report can be learned in a relatively short time and all are similar in psychological requirements as well as the equipment used. One cannot be sure that if different tasks were studied, different results might not be obtained. Nevertheless, as far as they go, the results are clear: no new factors emerge with practice. As it stands, this conclusion is liable to be confused with other, similar sounding but definitely different, propositions. Three such propositions need to be discussed specifically.

First, an absence of late-appearing factors does not mean that no factor grows stronger with practice. Fleishman and Rich (1963), for example, administered a spatial test and a measure of kinaesthetic sensitivity to 40 college students, who were then given 40 1-min trials on the Two-Hand Coordination

Test (THC). The measure of kinaesthetic sensitivity correlated more and more strongly with THC as practice progressed, beginning with a value of .03 for the first four trials and ending with a value of .40 for the last four trials. Kinaesthetic sensitivity clearly relates more strongly to THC late than early in practice. It is not, however, a new factor. The hypothesis under test is that factors appear with practice that could not be identified earlier on. In the Fleishman and Rich study kinaesthetic sensitivity was measured before practice began. Conceivably, a test that either involved or presupposed extensive practice might measure a late-appearing factor. To prove the point, however, one would have to show that it did, in fact, emerge with practice in the sense that we are using the word "emerge." One would have to show that the factor loaded heavily only after extensive practice on a specified set of tasks.

The second proposition concerns "specific" variance. If two tasks are both practiced, the correlation between them late in practice represents "common" variance; all other variance in the two tasks is specific. If more than two tasks are practiced, as in our case, the communality for trials late in practice represents common variance, and what remains is specific. Except for extreme cases, the hypothesis under test carries no implications for specific variance. Late-appearing factors involve common variance. Hence, the appearance of factors late in practice is incompatible with total specificity. To the extent, however, that any common variance exists late in practice, so may late-appearing factors. It is generally agreed, for example, that gross motor performance after extended practice is largely task specific (Henry & Nelson, 1956; Lindeburg, 1949). Nevertheless, between-task correlations late in practice frequently range as high as .40 or.50, certainly high enough to accommodate late-appearing factors if any exist. Of course, the absence of late-appearing factors (the conclusion we have reached) is fully compatible with high, even total task specificity.

The third proposition is more general, namely, that the absence of late-appearing factors does not exclude differential change with practice; and in our two studies both within- and between-task correlations definitely do change with practice. In Experiment 2, for example, all five tasks show "superdiagonal form" (Jones, 1969); that is, the smallest correlation is in the upper right-hand corner of the matrix (between the first and fourth stages of practice), while the largest correlations are in the superdiagonal (between neighboring stages). In both experiments Breakout correlates more strongly with ACM late than early in practice; that is, Breakout "converges" on ACM (Jones, Kennedy, & Bittner, Note 3). In Experiment 2 Race Car and Anti-aircraft converge on each other. The question then arises as to how these two things can be reconciled. How is it possible for the correlations between and within tasks to change if no new factors emerge with practice?

There are three major answers to this question. The first is that error components in the practiced tasks may change with practice. Suppose, for example, that they become smaller; all reliable components remain the same but error variance shrinks. In such a case all correlations between tasks will increase because all correlations will become less attenuated with practice. Similarly, of course, an increase in error variance (a reduction in reliability) would mean smaller correlations late than early in practice.

The second possibility is that different abilities (different factors) are tapped as practice proceeds. This is the usual interpretation of changing correlations with practice, the one adopted by Fleishman and Rich in their study of kinaesthetic sensitivity and by Fleishman and other co-workers in numerous other studies (for example, Fleishman, 1960; Fleishman & Hempel, 1954; Fleishman & Parker, 1959). In all these studies the abilities at issue are identified and measured <u>before</u> practice begins. Subjects utilize different abilities at different stages of practice, hence, task correlations change, but nothing new emerges. A task simply taps a different mix of abilities as practice progresses.

The third possibility is that the same abilities are involved at different stages of practice but the subjects change in relation to them. In factor-analytic terms it is the factor scores that change, not the factors themselves. As practice proceeds, subjects move to new positions on some factors (abilities); in varying degrees they improve and, as they do, the correlations between tasks change. Theoretically this third possibility is quite distinct from the second; it seems competent to the facts and has its adherents (Alvares and Hulin, 1972, 1973; Dunham, 1974). Unfortunately, as these authors have themselves pointed out, a general empirical test that would discriminate between the changing-task (Fleishman) and changing-subject models is difficult to imagine; in any case, no one has.

What matters for the present paper is that the existence or nonexistence of late-appearing factors is a different question than any previously raised in this field. Changing correlations with practice can be and have been explained in several ways without resort to late-appearing factors. If late-appearing factors could be demonstrated, it would certainly argue in favor of the changing-task conception, though in a sense never intended by any of its proponents to date. If, on the other hand, late-appearing factors do not exist, the fact in no way gainsays the changing task model; it does not gainsay the changing-subject model either. The existence or nonexistence of late-appearing factors is an independent question.

Finally, two technical points need to be made. In all our tasks a single global measure of performance was used. If late-appearing factors exist, one would like to measure them directly. That would mean using other, more specific measures. As long, however, as the question at issue is the existence of late-appearing factors a battery of psychologically more unitary measures is premature. The central claim of factor analysis is that it can detect latent (unmeasured) variations; and the design we have used allows us to conclude whether or not late-appearing latent variations exist. If the findings had favored the existence of such variations, the next step would have been to try to measure them directly. As it is, however, any such attempt is pointless.

In the factor analyses we carried out factoring was continued until all remaining factors had eigenvalues of less than unity. This is the conventional procedure. Nevertheless, it leaves open the possibility that some factors with small eigenvalues might load most heavily in late practice only. They could hardly, however, load heavily in an absolute sense at any point. Even if such factors existed, they could not be identified in the usual sense of the word. One could not, for example, calculate factor scores that would serve, even roughly, as a surrogate for the factor itself.

From a factor-analytic point of view the results of the present study are encouraging. If practice brought with it new factors, those factors could not be measured by the usual brief tests. In order to get at them one would have to provide extended practice, an expensive and laborious process. A failure, therefore, to find late-appearing factors is consonant with existing procedures.

From a training point of view, the same results are not so encouraging. The army is currently experimenting with video games as part-task trainers (Trachtman, 1981). But if practice does not produce new factors, what is the basis for transfer of training to operational tasks? In the absence of new factors, the only possibility is the third interpretation of differential change with practice mentioned earlier. That is, if transfer of training takes place, it must be that subjects are improving their scores along lines (abilities, factors) that could have been identified before practice began. Presumably their scores at that point would be lower. It may be, too, that these abilities can only be measured through video-computer as contrasted with more conventional tasks. In this sense, the current wave of video-computer tasks may uncover new dimensions of human skills and abilities; but practicing the tasks seems not to produce any further novelties.

FOOTNOTES

[1]Data collection for this study began in the fall of 1980 when fully programmable video systems for home television sets had just arrived on the market and were still relatively unfamiliar.

[2]Both left and right difficulty switches were set in the A position for all five games in Experiment 2.

REFERENCE NOTES

1. Jones, M.D. Further studies on stabilization with practice in performance tests. Final Report, Contract No. N00203-80-M-5387 with the Naval Biodynamics Laboratory, New Orleans, Louisiana, May 15, 1981, pp. 7-9.

2. Jones, M.B. Differential retention and convergence with practice. First Quarterly Report, Contract No. MDA903-81-C-0293 with the Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia, July 27, 1981, p. 14.

3. Jones, M.B., Kennedy, R.S., & Bittner, A.C., Jr. Video games and convergence or divergence with practice. Proceedings of the 7th Psychology in the Department of Defense Symposium, Colorado Springs, Colorado, April 1980.

REFERENCES

Alvares, K.M., & Hulin, C.L. Two explanations of temporal changes in ability-skill relationships: A literature review and theoretical analysis. Human Factors, 1972, 14, 295-308.

Alvares, K.M., & Hulin, C.L. An experimental evaluation of a temporal decay in the prediction of performance. Organizational Behavior and Human Performance, 1973, 9, 169-185.

Dreyfus, S.E., & Dreyfus, H.L. The scope, limits, and training implications of three models of aircraft pilot emergency response behavior. Operations Research Center, University of California (ORC79-2), February, 1979.

Dunham, R.B. Ability-skill relationships: An empirical explanation of change over time. Organizational Behavior and Human Performance, 1974, 12, 372-382.

Duncanson, J.P. Intelligence and the ability to learn. Princeton, N.J.: Educational Testing Service, 1964.

Dunlap, W.P., Jones, M.B. & Bittner, A.C., Jr. Average correlations versus correlated averages. Submitted to The British Journal of Mathematical and Statistical Psychology.

Fleishman, E.A. Abilities at different stages of practice in rotary pursuit performance. Journal of Experimental Psychology, 1960, 60, 162-171.

Fleishman, E.A., & Hempel, W.E., Jr. changes in factor structure of a complex psychomotor test as a function of practice. Psychometrika, 1954, 18, 239-252.

Fleishman, E.A., & Parker, J.F., Jr. Prediction of advanced levels of proficiency in a complex tracking task. USAFWAD Technical Report, 1959, No. 59-255.

Fleishman, E.A., and Rich, S. Role of kinesthetic and spatial visual abilities in perceptual-motor learning. Journal of Experimental Psychology, 1963, 66, 6-11.

Gundlach, R. The effects of practice on the correlations of three mental tests. Journal of Educational Psychology, 1926, 17, 387-401.

Henry, F.M., & Nelson, G.A. Age differences and inter-relationships between skill learning in gross motor performance of ten and fifteen year old boys. Research Quarterly, 1956, 27, 162-175.

Horn, J.L. State, trait, and change dimensions of intelligence. British Journal of Educational Psychology, 1972, 42 (part 2), 159-185.

Jarrett, J.W. Brain/2 =? Leading Edge, Summer 1979, 9-10.

Jones, M.B. Differential processes in acquisition. In E.A. Bilodeau and Ina McD. Bilodeau (Eds.), Principles of skill acquisition. New York: Academic Press, 1969.

Jones, M.B., Kennedy, R.S., & Bittner, A.C., Jr. A video game for performance testing. American Journal of Psychology, 1981, 94, 145-152.

Kennedy, R.S., Bittner, A.C., Jr., Harbeson, M., & Jones, M.B. Television computer games: A 'new look' in performance testing. Aviation, Space, and Environmental Medicine, 1982, 53, 49-53.

Lewis, D., McAllister, D.E., & Bechtoldt, H.P. Correlational study of performance during successive phases of practice on the standard and reversed tasks on the SAM complex coordinator. Journal of Psychology, 1953, 36, 111-126.

Lindeburg, F.A. A study of the degree of transfer between quickening exercises and other coordinated movements. Research Quarterly, 1949, 20, 180-194.

Rebert, C.S., & Low, D.W. Differential hemispheric activation during complex visuomotor performance. Electroencephologoraphy and Clinical Neurophysiology, 1978, 44, 724-734.

Schneider, W., & Shiffrin, R.M. Controlled and automatic human information processing; 1. Detection, search, and attention. Psychological Review, 1977, 84, 1-66.

Stake, R.F. Learning parameters, aptitudes, and achievements. Psychometric Monographs, 1961, No. 9.

Thomas, D.J., Majewski, P.L., Ewing, C.L., & Gilbert, N.S. Medical qualification procedures for hazardous duty aeromedical research. (Conference Proceedings No. 231, A3, pp. 1-13) London: AGARD, 1978.

Trachtman, P. A generation meets computer — and they are friendly. Smithsonian, 1981 (September), 12, 50-61.

Woodrow, H.A. The ability to learn. Psychological Review, 1946, 53, 147-158.

TABLE 1

Information regarding the five video games.

NAME	CASSETTE, GAME	DIFFI-CULTY	TASK DESCRIPTION	CONTROL DEVICE	TRIAL LENGTH	TRIALS/ DAY	Score
I) Air Combat Maneuvering	Combat (CX-2601), Game 24	Left=B Right=B	Attack opposing jet and shoot it	joystick, button	2 min. 16 sec.	10 trials/day 15 days	# of times opposing jet is hit
II) Breakout	Breakout (CX-2622), Game 1	Left=B Right=B	Ball ricochets from brick wall & is returned by paddle	paddle button	5 balls	10 trials/day 15 days	# of bricks knocked down (each "wall" of bricks is worth differet # points)
III) Surround	Surround (CX-2641), Game 4	Left=A Right=A	Computer controlled cursor must be circumnavigated	joystick	1st to score 10	5 trials/day 15 days	transformed ratio of computer: Ss # of surrounds
IV) Race Car	Indy 500 (CX-2511),	Left=A Right=A	Car must drive around course & attempt to avoid hitting wall (hitting wall slows car down)	special knob	60 sec.	15 trials/day 15 days	# of laps completed around the course
V) Slalom	Street Racer (CX-2612), Game 7	Left=A Right=A	Skier must pass through gates	paddle, button	2 min. 16 sec.	7 trials/day 15 days	# of gates passed through minus # of times skier hits a wall

TABLE 2

Correlations among five microcomputer-television tasks, early, midway, and late in practice, Experiment 1.

No.	Task	Stage of Practice	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	ACM	Early	–	.80	.82	.48	.64	.66	.49	.41	.40	.46	.26	.12	.21	.38	.30
2		Midway		–	.91	.46	.57	.57	.41	.32	.31	.36	.09	.05	.22	.41	.29
3		Late			–	.55	.64	.65	.44	.43	.40	.34	.17	.10	.21	.46	.35
4	Breakout	Early				–	.78	.73	.43	.40	.37	.23	.43	.22	.00	.43	.20
5		Midway					–	.82	.59	.50	.48	.43	.44	.21	.08	.43	.33
6		Late						–	.61	.58	.54	.50	.56	.32	.13	.50	.37
7	Race Car	Early							–	.65	.65	.53	.46	.40	.20	.46	.45
8		Midway								–	.72	.37	.48	.43	.31	.45	.47
9		Late									–	.35	.42	.44	.24	.42	.49
10	Slalom	Early										–	.43	.28	.19	.27	.25
11		Midway											–	.57	.18	.32	.10
12		Late												–	.31	.39	.16
13	Surround	Early													–	.31	.21
14		Midway														–	.44
15		Late															–

TABLE 3

Principal factors for the correlations in Table 2.

Task	Practice	Factor 1	Factor 2	Factor 3
ACM	Early	.76	-.39	.08
	Midway	.71	-.57*	.20
	Late	.79*	-.51*	.16
Breakout	Early	.69	0.11	-.45*
	Midway	.83*	-.14	-.34*
	Late	.88*	-.03	-.28
Race Car	Early	.75	.24	.08
	Midway	.72	.34	.18
	Late	.69	.33	.18
Slalom	Early	.54	.10	.01
	Midway	.55	.46	-.28
	Late	.42	.52*	.03
Surround	Early	.28	.16	.33*
	Midway	.60	.10	.14
	Late	.48	.09	.30

*The three largest correlations for each factor are marked with an asterisk.

TABLE 4

Varimax-rotated factors for the correlations in Table 2.

Task	Stage of Practice	Factor 1	Factor 2	Factor 3
ACM	Early	.76*	.29	.25
	Midway	.91*	.12	.18
	Late	.90*	.22	.23
Breakout	Early	.40	.72*	.06
	Midway	.52	.72*	.18
	Late	.48	.73*	.31
Race Car	Early	.28	.42	.61*
	Midway	.20	.35	.71*
	Late	.19	.33	.68*
Slalom	Early	.24	.33	.37
	Midway	-.09	.64	.41
	Late	-.15	.33	.56
Surround	Early	.12	-.06	.45
	Midway	.31	.26	.48
	Late	.28	.06	.50

TABLE 5

Quartimax-rotated factors for the correlations in Table 2.

Task	Stage of Practice	Factor 1	Factor 2	Factor 3
ACM	Early	.62*	.12	.57
	Midway	.80*	.03	.48
	Late	.76*	.08	.57
Breakout	Early	.23	.58*	.65
	Midway	.32	.51*	.68
	Late	.24	.44*	.77*
Race Car	Early	.04	.02	.79*
	Midway	-.04	-.09	.81*
	Late	-.04	-.10	.77*
Slalom	Early	.08	.07	.54
	Midway	-.31	.30	.63
	Late	-.33	-.03	.57
Surround	Early	.02	-.30	.36
	Midway	.13	-.04	.61
	Late	.13	-.22	.52

*The three largest correlations for each factor are marked with an asterisk. Four correlations with Factor 3 are marked because of the tie for third largest.

TABLE 6

Direct oblimin (oblique) factor structure for the correlations in Table 2.

Task	Stage of Practice	Factor 1	Factor 2	Factor 3
ACM	Early	.79*	.48	.46
	Midway	.92*	.34	.38
	Late	.92*	.44	.45
Breakout	Early	.42	.79*	.31
	Midway	.56	.83*	.45
	Late	.52	.86*	.57
Race Car	Early	.33	.57	.74*
	Midway	.26	.51	.80*
	Late	.25	.47	.77*
Slalom	Early	.28	.43	.48
	Midway	-.04	.67	.53
	Late	-.10	.38	.58
Surround	Early	.14	.04	.43
	Midway	.35	.40	.58
	Late	.31	.21	.55

*The three largest correlations for each factor are marked with an asterisk.

TABLE 7

Trial means and standard deviations of five microcomputer-television tasks, Experiment 2.

Task		Day 1	2	3	4	5	6	7	8	9	10	11	12
ACM	$\bar{x}$	6.9	8.2	10.0	11.0	12.2	13.6	13.9	14.6	15.3	16.1	16.2	16.6
	SD	2.3	2.9	3.1	3.0	3.1	3.6	3.2	3.3	4.0	3.8	3.5	3.3
Breakout	$\bar{x}$	345	435	490	529	577	623	654	706	752	771	854	860
	SD	177	197	195	210	211	252	259	286	285	274	283	257
Race Car	$\bar{x}$	7.1	8.9	9.7	10.3	10.8	11.1	11.4	11.6	11.8	12.0	12.2	12.2
	SD	1.7	1.7	1.4	1.3	1.2	1.2	1.1	1.0	1.0	1.0	0.9	1.0
Slalom	$\bar{x}$	41.3	52.4	57.3	59.9	64.0	69.0	68.8	71.9	73.0	73.1	73.1	77.2
	SD	11.3	12.4	13.0	13.8	13.7	13.9	12.6	13.3	14.7	14.4	15.2	13.1
Anti-aircraft	$\bar{x}$	14.4	18.1	20.4	22.5	23.8	25.7	26.6	27.2	28.0	29.1	29.3	30.3
	SD	4.1	5.3	5.3	5.3	5.5	5.7	5.8	6.0	5.8	5.3	5.8	5.9

TABLE 8

Correlations among five microcomputer-television tasks at four stages of practice, Experiment 2.

No.	Task	Stage of Practice	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	ACM	1		.67	.52	.50	.46	.45	.48	.50	.41	.37	.45	.53	.38	.34	.29	.28	.36	.44	.45	.49
2		2			.71	.63	.38	.42	.44	.45	.31	.20	.30	.39	.32	.33	.34	.32	.27	.40	.41	.47
3		3				.70	.33	.40	.44	.44	.30	.23	.32	.39	.35	.33	.37	.35	.26	.34	.41	.46
4		4					.33	.45	.40	.47	.29	.23	.33	.41	.37	.36	.37	.37	.25	.35	.41	.46
5	Breakout	1						.58	.57	.52	.47	.44	.48	.49	.39	.39	.26	.24	.33	.42	.38	.42
6		2							.57	.58	.38	.37	.45	.48	.40	.45	.34	.34	.26	.36	.37	.44
7		3								.62	.44	.46	.50	.52	.38	.31	.23	.20	.38	.42	.41	.46
8		4									.33	.41	.46	.50	.33	.37	.32	.29	.36	.43	.41	.47
9	Race Car	1										.50	.48	.44	.33	.25	.14	.10	.31	.31	.31	.37
10		2											.63	.59	.31	.31	.19	.19	.40	.38	.39	.42
11		3												.74	.40	.42	.33	.30	.45	.48	.52	.54
12		4													.39	.42	.31	.33	.49	.57	.61	.63
13	Slalom	1														.57	.49	.44	.23	.41	.31	.36
14		2															.66	.57	.24	.39	.26	.42
15		3																.65	.13	.35	.21	.37
16		4																	.18	.31	.26	.36
17	Antiaircraft	1																		.55	.51	.53
18		2																			.61	.63
19		3																				.68
20		4																				

TABLE 9

Principal factors for the correlations in Table 8.

Task	Stage of Practice	Factor 1	2	3	4
ACM	1	.70*	-.01	-.25	.03
	2	.66	.22	-.50*	-.04
	3	.65	.26	-.44*	-.04
	4	.64	.25	-.36*	-.03
Breakout	1	.66	-.14	.04	.31*
	2	.67	.03	-.01	.29
	3	.69	-.18	-.10	.27
	4	.69	-.05	-.09	.19
Race Car	1	.54	-.24	.00	.22
	2	.60	-.34*	.21	.14
	3	.72*	-.25	.24	.03
	4	.77*	-.24	.15	-.09
Slalom	1	.58	.23	.20	.09
	2	.61	.40*	.36*	.07
	3	.53	.58*	.31	-.04
	4	.50	.49*	.25	-.09
Antiaircraft	1	.54	-.30	.06	-.25
	2	.68	-.12	.10	-.30*
	3	.68	-.23	-.04	-.38*
	4	.75*	-.11	.02	-.32*

*The four largest correlations for each factor are marked with an asterisk.

TABLE 10

Varimax-rotated factors for the correlations in Table 8.

Task	Stage of Practice	Factor 1	Factor 2	Factor 3	Factor 4
ACM	1	.52*	.40	.15	.31
	2	.79*	.20	.17	.19
	3	.75*	.18	.22	.19
	4	.69*	.19	.26	.19
Breakout	1	.22	.65*	.19	.20
	2	.32	.58*	.29	.13
	3	.33	.64*	.10	.24
	4	.38	.53	.20	.24
Race Car	1	.16	.56	.05	.23
	2	-.01	.60*	.11	.40
	3	.06	.56	.25	.51
	4	.17	.50	.23	.60
Slalom	1	.18	.32	.52*	.17
	2	.13	.28	.74*	.16
	3	.20	.08	.81*	.11
	4	.20	.05	.70*	.16
Antiaircraft	1	.12	.26	.05	.61*
	2	.21	.24	.26	.64*
	3	.29	.21	.10	.72*
	4	.32	.25	.24	.67*

*The four largest correlations for each factor are marked with an asterisk.

TABLE 11

Quartimax-rotated factors for the correlations in Table 8.

Task	Stage of Practice	Factor 1	Factor 2	Factor 3	Factor 4
ACM	1	.34*	.66	.04	.00
	2	.66*	.54	.10	.01
	3	.63*	.52	.15	.02
	4	.56*	.52	.18	.01
Breakout	1	-.01	.70	.04	-.23
	2	.11	.66	.16	-.25
	3	.10	.74*	-.05	-.19
	4	.16	.68	.06	-.13
Race Car	1	-.04	.61	-.08	-.14
	2	-.24	.68	-.04	-.03
	3	-.19	.77*	.09	.08
	4	-.08	.80*	.07	.19
Slalom	1	.03	.50	.43*	-.06
	2	-.01	.49	.66*	-.06
	3	.11	.34	.77*	.01
	4	.12	.33	.66*	.07
Antiaircraft	1	-.06	.58	-.06	.34*
	2	.02	.65	.13	.37*
	3	.10	.66	-.02	.46*
	4	.12	.71*	.13	.38*

*The four largest correlations for each factor are marked with an asterisk.

TABLE 12

Direct oblimin (oblique) factor structure for the correlations in Table 8.

Task	Stage of Practice	Factor 1	Factor 2	Factor 3	Factor 4
ACM	1	.63*	.57	.37	.53
	2	.85*	.42	.39	.43
	3	.82*	.40	.43	.42
	4	.76*	.41	.45	.42
Breakout	1	.37	.73*	.38	.46
	2	.47	.68	.47	.41
	3	.46	.74*	.32	.50
	4	.51	.66	.40	.49
Race Car	1	.27	.62	.22	.43
	2	.15	.68	.28	.57
	3	.25	.71*	.43	.69
	4	.37	.69*	.44	.78*
Slalom	1	.34	.46	.62*	.38
	2	.33	.44	.81*	.37
	3	.37	.26	.84*	.29
	4	.36	.24	.74*	.31
Antiaircraft	1	.24	.44	.22	.67
	2	.37	.48	.43	.74*
	3	.43	.46	.31	.80*
	4	.48	.52	.46	.80*

*The four largest correlations for each factor are marked with an asterisk.

Summary and Conclusions

Widespread availability of high speed, large storage, economic computers has set the stage for dramatic change in the assessment of differential abilities. It appears clear that assessment procedures controlled, administered, and scored by computer will shortly become a viable alternative to paper-and-pencil tests. Computer testing opens also the prospect of assessing capacities not measurable with paper-and-pencil: in particular, computer testing allows the assessment of motor skills on tasks modeled after video games. The present research involved administering extended practice on five commercial video games, as well as a battery of paper-and-pencil tests. The research questions were: 1) Does the factorial content of the video games change with practice and, if so, is there a late appearing factor(s) that ties in with an hypothesized shift from controlled to automatic processing with practice ? 2) Do video games involve new common factors which have not been and perhaps cannot be identified in paper-and-pencil tests ? 3) To what extent can video-game performance be predicted from conventional paper-and-pencil tests ?

Findings - Part 1. Two methods of smoothing Pearson r's were investigated by a Monte Carlo computer simulation. One method was to average repeated measures in blocks and then to correlate block averages. In the other method all repeated measures were correlated and then the correlation coefficients were averaged in blocks, after converting via Fisher's z transformation to avoid bias. The latter approach proved much superior, resulting in greatly reduced sampling variability, and little distortion in the estimated population correlation. Therefore, if one's purpose is to describe the relation between learning trials, averaging correlations produces a much better result than correlating averages.

Findings - Part 2. Part 2 of this report examined the possibility of late appearing factors consequent to a switch from controlled to automatic processing. If differential factors associated with automatic processing develop, they could exist only after this switch and, hence, could be identified only late in practice. Two tests of this reasoning were carried out; both used extended practice on five video games, but in two different populations and approximately one year apart. The results of the experiments are in striking agreement. In both cases, with one possible exception in the second experiment, factors were identified by content exclusively and not by stage of practice. Other studies using other materials, subjects, or conditions of practice may reach different conclusions; however, the studies reported in this paper offer no support for the existence of late-appearing factors and may be interpreted to challenge the notion of an identifiable shift from controlled to automatic processing.

We are currently evaluating the relations of the one-time paper-and-pencil tests to the various stages of practice on the video games, with the aim of identifying and characterizing that portion of motor skills performance that is not predictable from the one-time tests. This would be the kind of information that skills tasks might add to a standard test battery.

Suggestions for Future Research. We envision three major directions for pursuing these studies. First, more information is needed on the relations between paper-and-pencil tests and video games. Our work represents only a beginning, as we used one battery of conventional tests out of a large number of possible test batteries. We also used only one set of video games and there are many others. Second, retention and learning studies are needed to determine the temporal stability of video-computer skills. How well are the test retained on the average, and how well does performance at the end of acquisition correlate with performance at retention? Both questions must be asked for varying lapses of time between acquisition and retention testing. Third, the games themselves need to be extended. There is no way to modify existing commercial video games to suit experimental purposes. Similarly, the scoring of a player's performance still must be recorded by hand. Close facsimilies to many of the commercial games are available on the Apple II or other personal computers, and a burgeoning market now exists for game programs written specifically for personal computers. These new developments should be exploited in the interests of greater experimental control and flexibility of scoring.

REPORT DISTRIBUTION LIST

Navy

1 Ahlers, Robert
1 Bachrach, Arthur
1 Baker, Meryl S.
1 Biersner, Robert
1 Bittner, Alvah
1 Bond, Nicholas A.
1 Bory, Alexander (NAMRL)
1 Collyer, Stan
1 CURRAN MIKE
1 Duffy, Tom
1 Englund, Carl E.
1 Federico, Pat-Anthony
1 FORD JOHN
1 Gaynor, Mike
1 HARRIS STEVE
1 HOLLAN JIM
1 KERR NORMAN J
1 Lester, James
1 MALOY WILLIAM L
1 Martin, Richard L.
1 MCBRIDE JAMES
1 Moeller, George
1 Montague, William
1 NPRDC LIBRARY
1 NPRDC TECHNICAL DIRECTOR
6 NRL CODE 2627
1 ONR 441NP
6 ONR 442PT
1 ONR PASADENA
1 OP115
1 Petho, Frank C.
1 POOCK GARY
1 Rimland, Bernie
1 Smith, Robert G.
1 SMODE ALFRED F
1 SOPENSEN RICHARD
1 Thomson, Gary
1 Wegman, Dr. Edward
1 WEISSINGER-BAYLON ROGER
1 WEITZMAN RONALD
1 Wetzel, Douglas
1 Woodward, Donald

Marine Corps

1 GREENUP H WILLIAM (MCDEC)
1 ONR 100M
1 SLAFKOSKY AL

Private Sector

1 Annett, John
1 Australian DoD Psychology
1 Baddeley, Alan
1 Bejar, Isaac
1 Bock, R. Darrel
1 BOHRER, ARNOLD
1 Bund. der Verteidigung
1 Carroll, John B.
1 CHASE WILLIAM
1 CROMBAG HANS
1 Damos, Diane
1 EGGENBERGER J C (CANADA)
1 ERIC
1 Feuerstein, Reuven
1 Green, Bert
1 HAWKINS HAROLD
1 HOFFMAN JAMES
1 HUMPHREYS LLOYD
1 Hunt, Earl
1 Jones, Douglas H.
1 KEELE STEVEN W
1 KELSO, SCOTT
1 Lansman, Marcy
1 LEVINE MICHAEL
1 LEWIS CHARLES
1 LINN ROBERT
1 LUMSDEN JAMES
1 NOVICK MELVIN R
1 NRC HUMAN FACTORS COMMITTEE
1 Orlansky, Jesse
1 PAULSON JAMES A
1 Petrullo, Luigi
1 Poltrock, Steven
1 ROSE ANDREW M
1 RUDNER LAWRENCE
1 Schn ider, Walter
1 SNOW RICHARD
1 Spoehr, Kathryn T.
1 Sternberg, Robert
1 Tatsuoka, Kikumi
1 Vale, David
1 Weiss, David J.
1 WHITELY SUSAN

Army

1 ARI Technical Director
1 Baker, James
1 FARR BEATRICE
1 KATZ MILTON S
1 SASMOR ROBERT
1 Wing, Hilda
1 WISHER ROBERT

Air Force

1 AFHRL/LRS Library
1 AFHRL MPD
1 AFOSR Life Sciences
1 Air University Library
1 Alluisi, Earl
1 Christal, Raymond E.
1 Fregly, Alfred
1 HADDAD GENEVIEVE
1 Hunter, David
1 Longridge, Tom
1 Yasatuke, Joe

Department of Defense

12 DTIC
1 OUSDRE
1 Thorpe, Jack A.

Civilian Agencies

1 CHIPMAN SUSAN
1 URRY VERN W
1 USCG PSYCHOLOGICAL RESEARCH
1 WARM THOMAS
1 Young, Joseph L.

DATE
ILME